

Università di Cassino
Corso di Statistica 1
Esercitazione del 10/10/2006
Dott. Alfonso Piscitelli

Esercizio 1

Supponiamo che il collettivo che si vuole studiare sia composto da un gruppo di amici. La seguente tabella raccoglie alcuni dati su di essi.

- a. Qual è l'unità statistica? Quante unità statistiche sono presenti?
- b. Quali sono i caratteri? Per ogni carattere indicare se è quantitativo, discreto o continuo, su scala ad intervalli o su scala di rapporti, oppure qualitativo sconnesso o ordinato.
- c. Costruire le distribuzioni di frequenza per i caratteri **Colore degli occhi**, **Sesso** e **Nucleo familiare**.
- d. Costruire la distribuzione di frequenza per il carattere **Età** suddividendo la distribuzione in 6 classi equiampie.
- e. Rappresentare graficamente la distribuzione equiampia.
- f. Costruire la distribuzione di frequenza per il carattere **Reddito** suddividendo la distribuzione in 5 classi equifrequenti.
- g. Rappresentare graficamente la distribuzione equifrequente.

NOME	Sesso	Età	Reddito (Migliaia di €)	Titolo di studio	Nucleo familiare	Statura (cm)	Colore degli occhi
Sergio	M	22	0,7	Diploma	3	173	NERO
Teresa	F	18	0,2	Lic. Media	4	168	MARRONE
Emilia	F	34	1,6	Diploma	2	165	MARRONE
Giorgio	M	42	2,5	Laurea	5	180	NERO
Valeria	F	50	3,2	Diploma	3	163	AZZURRO
Claudia	F	12	0,1	Lic. Elementare	4	160	NERO
Gianfranco	M	46	3,8	Lic. Media	4	177	MARRONE
Giovanni	M	72	1,3	Nessun Titolo	2	164	VERDE
Monica	F	27	1,2	Laurea	3	158	AZZURRO
Simona	F	48	1,7	Lic. Media	5	170	NERO
Federica	F	35	1,9	Laurea	1	167	NERO
Alfonso	M	84	0,8	Nessun Titolo	1	159	MARRONE
Giovanna	F	21	0,4	Diploma	5	174	AZZURRO
Beatrice	F	44	1,8	Diploma	4	164	VERDE
Saverio	M	56	1,9	Lic. Media	2	177	NERO
Donatella	F	58	3,2	Lic. Media	3	172	NERO
Francesca	F	37	2,1	Diploma	1	166	MARRONE
Elena	F	16	0,1	Lic. Media	4	160	MARRONE
Massimo	M	73	1,6	Lic. Elementare	2	170	AZZURRO
Antonio	M	64	2,2	Lic. Elementare	3	184	VERDE

Soluzioni

- a) L'unità statistica è rappresentata da ogni individuo, appartenente al collettivo oggetto di studio, sul quale sono state registrate le informazioni oggetto dell'indagine. Sono presenti 20 unità statistiche.
- b) I caratteri sono l'oggetto dell'indagine, le *caratteristiche* che si intendono rilevare sul collettivo di individui. I caratteri possono essere **qualitativi** (sconnesso o ordinabile) o **quantitativi** (discreto o continuo).

Carattere	Tipo	Livello di misura	MODALITÀ-INTENSITÀ
Sesso	Qualitativa sconnessa	Scala Nominale	{Maschio, Femmina}
Età	Quantitativa continua	Scala di Rapporti	[12; 84]
Reddito	Quantitativa continua	Scala di Rapporti	[0,1; 3,8]
Titolo di studio	Qualitativa ordinabile	Scala Ordinale	{Nessun Titolo, Lic. Elementare, Lic. Media, Diploma, Laurea}
Nucleo familiare	Quantitativa discreta	Scala di Rapporti	[1; 5]
Statura	Quantitativa continua	Scala di Rapporti	[158; 184]
Colore occhi	Qualitativa sconnessa	Scala Nominale	{Nero, Marrone, Azzurro, Verde}

- c) La **frequenza assoluta** non è altro che il numero di volte che una determinata intensità o modalità si presenta nel collettivo statistico. In altre parole la frequenza n_i associata alla i -esima intensità (modalità) è uguale al numero di unità statistiche che presentano quella intensità (modalità).

La **distribuzione di frequenza semplice** associa alle modalità o intensità che può assumere un carattere X , le corrispondenti frequenze assolute. Per costruire una distribuzione di frequenza semplice partendo da una successione di osservazioni è sufficiente assegnare ad ogni intensità (o modalità, nel caso di variabili qualitative) la propria frequenza assoluta.

Le **frequenze relative** f_i , invece, sono date dal rapporto tra la frequenza assoluta, n_i , ed il totale delle osservazioni, N .

Quindi la distribuzione di frequenze del carattere (variabile) **Sesso** è:

Sesso	n_i	f_i
Maschi	8	0,4
Femmine	12	0,6

Tot: 20 1

Quella associata alla variabile **Colore degli occhi** è:

Colore degli occhi	n_i	f_i
Nero	7	0,35
Marrone	6	0,3
Azzurro	4	0,2
Verde	3	0,15

Tot: 20 1

Infine, la distribuzione di frequenze della variabile **Nucleo familiare** è:

Nucleo familiare	n_i	f_i
1	3	0,15
2	4	0,2
3	5	0,25
4	5	0,25
5	3	0,15

Tot: 20 1

d) Nella tabella sono riportati i valori della variabile **Età**.

Osservazione	Età
1	22
2	18
3	34
4	42
5	50
6	12
7	46
8	72
9	27
10	48
11	35
12	84
13	21
14	44
15	56
16	58
17	37
18	16
19	73
20	64

Al fine di pervenire a delle classi equiampie è necessario individuare il campo di variazione della variabile che si vuole suddividere in classi ed in seguito dividerlo per il numero di classi desiderate.

In questo caso, l'ampiezza delle classi è ottenuta come il rapporto tra il campo di variazione (il valore massimo della distribuzione – il valore minimo) e il numero di classi:

$$d = \frac{X_{max} - X_{min}}{K}$$

dove **K** = numero di classi.

Nel nostro caso: $x_{max} = 84$ $X_{min} = 12$ **K** = 6 \Rightarrow **d** = 12

Una volta individuato quanto devono essere ampie le classi si può procedere alla costruzione della distribuzione per classi.

Partendo dal valore minimo (X_{min}) si aggiunge l'ampiezza di classe

Età	d_i	n_i	f_i	F_i	h_i
12- 24	12	5	0,25	0,25	0,42
24- 36	12	3	0,15	0,4	0,25
36- 48	12	4	0,2	0,6	0,33
48- 60	12	4	0,2	0,8	0,33
60- 72	12	1	0,05	0,85	0,08
72- 84	12	3	0,15	1	0,25
Tot:		20	1		

Dove:

n_i = frequenze assolute

f_i = frequenze relative

F_i = frequenze cumulate relative

h_i = densità di frequenza, cioè il rapporto tra la frequenza della i-esima

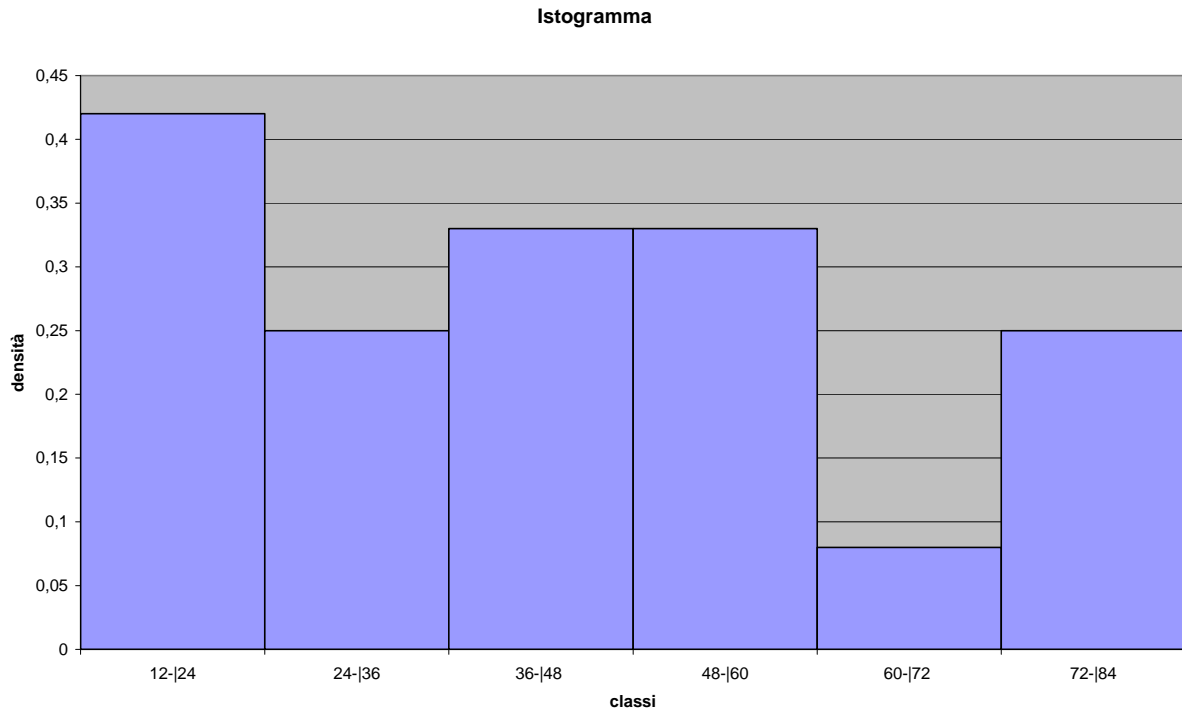
classe e la sua ampiezza: $h_i = \frac{f_i}{d_i}$.

Si precisa che nel caso di classi equiampie l'informazione ottenuta dalla densità di frequenza è la stessa che avremmo analizzando la distribuzione di frequenze semplice.

e) Sulla base dei risultati ottenuti nella precedente tabella è possibile costruire l'**istogramma** associato alla nostra distribuzione di frequenze.

L'istogramma normalizzato presenta sulle ascisse gli estremi delle classi e sulle ordinate la densità di frequenza. Essendo le classi di ampiezza uguale

le basi dei rettangoli associati alle diverse classi avranno tutte la stessa lunghezza.



f) Nella tabella sono riportati i valori ordinati della variabile **Reddito**.

Osservazione	Reddito
6	0,1
18	0,1
2	0,2
13	0,4
1	0,7
12	0,8
9	1,2
8	1,3
3	1,6
19	1,6
10	1,7
14	1,8
11	1,9
15	1,9
17	2,1
20	2,2
4	2,5
5	3,2
16	3,2
7	3,8



Avendo a disposizione 20 unità e dovendo costruire 5 classi equifrequenti (cioè con la stessa frequenza) viene da se che ogni classe dovrà avere una frequenza pari a $20 / 5 = 4$. In altre parole ogni classe deve contenere 4 unità statistiche.

La prima classe sarà quindi ottenuta prendendo i primi 4 valori della variabile in esame, la seconda i successivi 4 e così via.

Nel nostro caso:

Reddito	d_i	n_i	f_i	F_i	h_i
0,1 - 0,4	0,3	4	0,02	0,02	13,3
0,4 - 1,3	0,9	4	0,02	0,04	4,4
1,3 - 1,8	0,5	4	0,02	0,06	8,0
1,8 - 2,2	0,4	4	0,02	0,08	10,0
2,2 - 3,8	1,6	4	0,02	1	2,5
Tot:		20	1		

- g)** Sulla base dei risultati ottenuti nella precedente tabella è possibile costruire l'**istogramma** associato alla nostra distribuzione di frequenze. Essendo le classi di ampiezza diversa le basi dei rettangoli associati alle diverse classi avranno lunghezza diversa, in particolare maggiore è l'ampiezza della classe più deve essere lunga la base del rettangolo associato a quella classe.

